全原子力场的最新发展与天然无序蛋白质的模拟

陈家楠², 吴云东^{1,2,3*}

1. 深圳湾实验室, 深圳 518055

2. 北京大学深圳研究生院, 深圳 518055

3. 北京大学化学与分子工程学院, 北京 100871

*通讯作者, E-mail: wuyd@pku.edu.cn

收稿日期: 2020-07-21; 接受日期: 2020-08-20; 网络版发表日期: 2020-09-28

摘要 天然无序蛋白质是一类特殊的蛋白质, 在天然条件下不会折叠成特定的结构, 却有着重要的生物学功能. 分子动力学模拟是研究天然无序蛋白质的强大工具, 能够提供实验无法观测到的结构和动力学信息. 本文先总结了近年来全原子力场的发展以及不同力场在天然无序蛋白质体系中的表现, 然后概述了水模型对蛋白质模拟的影响, 最后介绍了分子动力学模拟在天然无序蛋白质结合过程研究上的应用.

关键词 分子动力学模拟, 蛋白质分子力场, 水模型, 天然无序蛋白质

1 引言

很多蛋白质都需要折叠成正确的结构来发挥功能, 但是有一类蛋白质无法形成特定的结构, 称为天然无序蛋白质(intrinsically disordered proteins, IDPs)^[1]. IDPs不能自发折叠成一个稳定的构象, 它们往往在很多不同的构象之间快速转变. IDPs通常含有较多的极性和带电氨基酸残基, 缺少非极性氨基酸残基来形成稳定的疏水核心. IDPs是完全无序的, 另外一些蛋白质则同时含有折叠的区域和有序的区域, 其中无序的区域被称为天然无序区域(intrinsically disordered regions, IDRs). 在本文中, IDPs和IDRs统称为IDPs. IDPs在真核基因组中非常常见, 44%的人类蛋白质编码基因含有超过30个氨基酸残基的无序片段^[2]. 在许多生物学过程中, IDPs起着至关重要的作用, 例如, 参与细胞信号传导, 调控转录、翻译、细胞周期等^[3~5].

因为IDPs的高度无序, 一些常用的结构生物学实验方法, 如X射线衍射和冷冻电子显微镜无法用于研究IDPs. 常见的研究IDPs的实验方法包括小角度X射线散射(small angle X-ray scattering, SAXS)^[6]、荧光共振能量转移(fluorescence resonance energy transfer, FRET)^[7]和核磁共振(nuclear magnetic resonance, NMR)^[8]. SAXS和FRET能提供分子内和分子间距离分布的信息, NMR能提供局部和长程的结构信息. 但是这些实验数据是IDPs复杂的构象系综的平均值, 无法直接从这些实验数据推测出IDPs的构象系综. 蛋白质力场是用于描述蛋白质势能的函数, 结合分子动力学(molecular dynamics, MD)模拟, 可以产生IDPs的构象系综, 提供实验无法得到的IDPs的结构和动力学信息. 此外, MD在研究IDPs的聚集^[9]、IDPs和生物大分子的结合^[10]以及IDPs和小分子的结合^[11,12]等方面也有广泛的应用(图1)^[13].

引用格式: Chen J, Wu Y. Recent development of atomistic force fields and simulations of intrinsically disordered proteins. *Sci Sin Chim*, 2020, 50: 1320–1332, doi: 10.1360/SSC-2020-0143

图 1 IDPs 折叠与结合的耦合^[13] (网络版彩图)
Figure 1 IDPs' coupled folding-binding process^[13] (color online).

本文主要关注的是全原子可加和蛋白质分子力场 (all-atom additive protein force field), 全原子可加和力场相比, 粗粒化力场和可极化力场能更好地平衡计算效率和准确度, 下文用力场指代全原子可加和蛋白质分子力场. 本文先总结了近年来力场的发展, 不同力场在 IDPs 体系中的测试, 然后概述了水模型对蛋白质模拟的影响, 最后介绍了分子动力学模拟在 IDPs 与 DNA、蛋白质以及小分子结合过程研究上的应用.

2 力场研究进展

蛋白质力场的发展始于 20 世纪 80 年代初, 在过去的 30 年间, 研究者们发展出多个系列的蛋白质力场, 主要包括 AMBER^[14]、CHARMM^[15]、GROMOS^[16] 和 OPLS^[17] 等, 一些综述对力场的发展进行了很好的总结^[18–21]. 本节介绍了近几年开发的力场, 包括 RSFF 系列力场^[22–24]、OPLSIDPSFF^[25]、ESFF1^[26]、ff19SB^[27]、Des-Amber^[28] 和 GROMOS 54A8_bb^[29]. 其中一些力场并不是针对 IDPs 优化的, 但是, IDPs 和有序蛋白质中存在的相互作用是相同的, 针对有序蛋白质的优化往往也可以改进 IDPs 的模拟结果.

修改氨基酸残基二面角参数来拟合蛋白质卷曲库^[30–32]中对应残基的二面角分布是近年来用的较多的一种力场优化策略^[33]. 对蛋白质卷曲库中不同残基的二面角分布进行统计可以得到各个残基内在构象的自由能分布, 通过调整各个残基的二面角参数可以将这些构象自由能整合到分子力场中. 通过对 OPLS-AA/L、ff99SB 和 ff14SB 力场进行残基特异性二面角

修正, Jiang、Zhou 和 Kang 等分别发展了 RSFF1^[22]、RSFF2^[23] 和 RSFF2C^[24] 残基特异性力场. 因为残基内在构象的自由能分布得到了较好的描述, 这些分子力场可以很好地折叠一系列迷你蛋白^[34,35]. Li 和 Es-cock^[36] 模拟了 256 个含有两个残基的多肽, 发现 RSFF2 得到的³ $J_{\text{NH}\alpha}$ 耦合常数和实验值的 R^2 达到了 0.82, 远远好于 ff99SB-ILDN^[37] 和 ff99SB-ILDN-NMR^[38] 的模拟结果 (图 2). 此外, RSFF 系列力场在蛋白质结构优化^[39]、IDPs 的模拟^[40,41] 和环状多肽的模拟^[42] 等方面也有很好的表现.

OPLSIDPSFF、ESFF1 和 Des-Amber 的发展也使用了上述的策略. Yang 等^[25] 基于 OPLS-AA/L 力场对骨架二面角进行了残基特异性修正得到新力场 OPLSIDPSFF. 对 2 个短肽、2 个有序蛋白质和 11 个 IDPs 的模拟表明, 二面角修正明显提高了力场的准确度, 模拟得到的化学位移、 J 耦合常数和回转半径 (radius of gyration, R_g) 和实验值符合得更好. 不仅如此, OPLSIDPSFF 搭配 disp-TIP4PD 水模型^[43] 相比 TIP3P 水模型^[44] 进一步提高了模拟结果, 在 A β 40、A β 42、ACTR、IA3、Tau 和 Mev-N 这些 IDPs 体系上尤为明显. ESFF1 和 OPLSIDPSFF 一样修改残基骨架二面角参数来拟合卷曲库中对应残基的骨架二面角分布, 但是 ESFF1 还考虑了邻位效应, 对同一种残基根据其序列上相邻残基的极性使用不同的骨架二面角参数. 对 61 个短肽和 IDPs 的测试表明, ESFF1 能够准确地重现化学位移和 J 耦合常数的实验值. ff19SB 是基于 ff14SB 优化得到的^[45], 同样对残基的骨架二面角参数进行了修正, 但是拟合的是隐式水模型中二肽的量子力学 (quantum mechanical, QM) 计算结果. ff19SB 结合 OPC 水模型^[46] 相比 ff14SB 结合 TIP3P 水模型显著提高了 Ace-A $_4$ XA $_4$ -NH $_2$ 体系中不同残基的螺旋倾向的模拟结果, 模拟值和实验值的 R^2 从 0.38 提高到了 0.75 (图 3). Piana 等^[28] 对 a99SB-disp 力场的非键和二面角参数进行全面修正得到新力场 DES-Amber. 他们先修改骨架原子的电荷和 Lennard-Jones (L-J) 相互作用参数来拟合 N -甲基乙酰胺和甘氨酸二肽的 QM 计算结果, 然后修改骨架二面角参数来拟合蛋白质卷曲库的二面角分布, 之后对骨架二面角参数又进行细微的调整来拟合不同氨基酸螺旋倾向的实验值. 对于侧链, 他们结合 QM 计算结果和蛋白质数据库中侧链旋转异构体的分布对侧链二面角参数进行修正, 并且对侧链和离子的

图2 256个含有两个残基的多肽的 $^3J_{\text{HNH}\alpha}$ 耦合常数模拟值和实验值的比较。(a) ff99SB-ILDN; (b) ff99SB-ILDN-NMR; (c) RSFF2的结果^[36] (网络版彩图)

Figure 2 Comparison of computed $^3J_{\text{HNH}\alpha}$ coupling constants with experimentally measured values for 256 two-residue peptides. Panels show the results for: ff99SB-ILDN (a), ff99SB-ILDN-NMR (b), and RSFF2 (c) ^[36] (color online).

非键参数进行修正来拟合渗透系数和Kirkwood-Buff积分的实验值。大量的测试证明, DES-Amber对于IDPs和有序蛋白质的模拟能够达到和a99SB-disp相当的程度, 并且能够稳定a99SB-disp无法稳定的蛋白-蛋白复合物, 更准确地预测结合自由能。Diem等^[29]发展了一个基于蒙特卡罗模拟和最速下降法的搜索算法来优化骨架二面角参数, 基于GROMOS 54A8力场^[47], 通过拟合二肽的 J 耦合常数和二级结构分布, 对骨架二面角的参数进行优化, 得到新力场GROMOS 54A8_bb。对52个蛋白质的测试表明, 新力场对于骨架二面角分布的描述更加准确。值得注意的是, 上述力场对不同氨基酸残基使用不同的骨架二面角参数, 只有GROMOS 54A8_bb对丙氨酸残基和甘氨酸残基以外的残基使用相同的骨架二面角参数。

3 力场测试

对不同力场用相同的体系进行测试, 能够客观地描述不同力场的表现, 并且为力场改进提供线索。本节对近年来用IDPs和有序蛋白质对不同力场进行测试比较的工作进行了总结(表1), IDPs体系包括 β -淀粉样多肽16~22片段($\text{A}\beta_{16-22}$)^[48,49]、肿瘤抑制因子p53的转录激活结构域(p53-TAD)^[50]和von Willebrand因子(VWF)的TIL'E'区域^[51], 折叠蛋白体系包括泛素、GB3蛋白^[52]和肌联蛋白I91结构域(titin I91)^[53]。下文对CHARMM力场和GROMOS力场名称使用缩写, 例如, CHARMM22*缩写为C22*, GROMOS 54a7缩写为G54a7。

IDPs相比有序蛋白质对力场参数更加敏感, 是非常好的用来测试力场的体系。过去用来测试的IDPs大多是较短的多肽, 如16个氨基酸的RS多肽^[54]、24个氨基酸的Histatin5^[55]。现在研究越来越多地使用更加复杂的体系, 复杂的体系对于力场模拟更具挑战, 也更容易暴露出力场的缺陷。例如, ff99SB-ILDN能够较好地模拟p53-TAD的40~61片段(p53-TAD2)^[56], 但是对于p53-TAD的1~61片段产生了过于塌缩的结构^[50]。Carballo-Pacheco等^[48]用G54a7、OPLS-AA、ff03ws、C22*和ff99SB*-ILDN 5个力场对 $\text{A}\beta$ 多肽的一部分, 7个氨基酸长度的 $\text{A}\beta_{16-22}$ 和它的两个突变体分别形成六聚体的过程进行了模拟。结果表明, 力场对低聚体形成的动力学性质有很大影响, 其中, G54a7和OPLS-AA形

图3 模拟得到的不同残基的螺旋倾向和实验值的相关性, 模拟用了4种力场和水模型的组合, 分别是(a) ff14SB+TIP3P、(b) ff14SB+OPC、(c) ff19SB+TIP3P、(d) ff19SB+OPC^[27] (网络版彩图)

Figure 3 Correlation between helical propensities from experiment and simulations using (a) ff14SB+TIP3P, (b) ff14SB+OPC, (c) ff19SB+TIP3P, and (d) ff19SB+OPC, respectively ^[27] (color online).

表1 测试力场的研究

Table 1 Force field benchmark studies

体系	实验数据	测试力场	参考文献
$\{\text{A}\beta_{16-22}\}_6$, $\{\text{A}\beta_{16-22}(\text{F19L})\}_6$, $\{\text{A}\beta_{16-22}(\text{F19V}, \text{F20V})\}_6$	解离常数	ff99SB*-ILDN, ff03ws, C22*, G54a7, OPLS-AA	[48]
$\{\text{A}\beta_{16-22}\}_2$		ff94, ff99, ff03, ff06, ff99SB-ILDN, ff12SB, ff14SB, C22*, C36, C36m, G43a1, G43a2, G45a3, G53a5, G5a6, G54a7, OPLS-AA	[49]
p53-TAD2	化学位移、残留偶极耦合	ff03, ff99SB-ILDN, C27, C36m, OPLS-AA/L	[56]
p53-TAD	化学位移、顺磁弛豫增强、smFRET和TR-FRET数据	ff99SB-ILDN, a99SB-disp, C22*, C36m	[50]
TIL'E' region of VWF	NOE效应、化学位移、NMR结构	ff99SB*-ILDN, ff14SB, a99SB-disp, RSFF2, RSFF2+, C27, C36m, C36m IDP	[51]
ubiquitin, GB3	3J 耦合常数、残留偶极耦合	ff03, ff03*, ff99SB-ILDN, ff99SB*-ILDN, ff14SB, AMBER FB15, C22, C22*, C27, C36, G54a7, OPLS-AA	[52]
titin I91		ff99SB-ILDN, ff99SB*-ILDN, ff14SB, AMBER FB15, C22*	[53]

成低聚体的速度最快, ff03ws^[57]最慢. 不同力场模拟得到的多肽内的接触模式比较类似, 而多肽间的接触模式有一定的差异, 说明不同力场低聚体形成速度的差异主要来自于多肽间相互作用. 不同力场模拟得到的接受碰撞概率也大不相同, 只有ff03ws的模拟结果符合预期, 并且能根据接受碰撞概率的大小区分能聚集的和不能聚集的 $\text{A}\beta_{16-22}$ 突变体. Man等^[49]用17个力场对 $\text{A}\beta_{16-22}$ 二聚体的结构和动力学性质进行了研究, 同样发现力场对纤维形成时间有很大影响. 进一步研究表明, 结构中 β -折叠的含量影响了纤维的形成时间, 结构中 β -折叠含量越高, 纤维形成时间越短. 他们对17个

力场进行分类和筛选, 去掉了过稳定 α -螺旋的力场ff94、ff99和ff12SB, 过稳定 β -折叠的力场ff96、G45a3、G53a5、G53a6、G43a1、G43a2和G54a7, 以及在其他研究中表现较差的力场ff03^[58]和OPLS-AA^[48], 认为剩下的5个力场ff99SB-ILDN、ff14SB、C22*、C36和C36m更加适合研究淀粉样肽聚集. Liu等^[50]用6个力场对p53-TAD的1~61片段进行了模拟, 将MD的模拟结果与NMR、单分子FRET(single-molecule FRET, smFRET)和时间分辨FRET(time-resolved FRET, TR-FRET)的实验数据进行了比较. 结果表明, C36m和C36mw没有收敛, C22*和ff99SB-ILDN产生的结构过

于紧凑, C22*还过稳定了螺旋构象, ff99SB-ILDN搭配 TIP4P-D后能够产生更加舒展的结构, 但是降低了蛋白-蛋白间相互作用的稳定性, 严重低估了蛋白质中螺旋的比例, 只有a99SB-disp能较好地重现所有的实验数据(图4(b)). VWF的TIL'E'区域(图4(a))同时包括折叠的E'结构域和几乎没有二级结构的TIL'结构域^[59], 对于力场模拟非常具有挑战. Kuzmanic等^[51]用8个力场对VWF的TIL'E'区域进行了模拟, 发现只有C36m、a99SB-disp和RSFF2+在TIL'结构域上形成了和NMR构象系综类似的较短的 β -折叠模体, 此外, ff99SB*-ILDN、C27、ff14SB和a99SB-disp会过稳定TIL'结构域上的 3_{10} 螺旋角. 综合比较残基间接触情况、NOE (nuclear overhauser effect)效应、化学位移、二级结构和骨架二面角, a99SB-disp和RSFF2+^[60]与实验值符合得最好.

不同力场对于有序蛋白质的模拟同样表现出一定的差异. Petrovic等^[52]用12个力场对泛素和GB3进行了模拟, 发现力场对氨基酸残基侧链旋转异构体的分布有很大影响, 其中, ff14SB和实验符合得最好. 此外, 比起蛋白质内部的残基, 力场对于蛋白质表面暴露的残基的侧链旋转异构体的分布模拟更不准确. 用原子力显微镜可以观察到, titin I91在力的作用下N端的10

个氨基酸会解折叠, 力弛豫后, 又会重新折叠^[61]. Wang等^[53]用5个AMBER力场和C22*^[62]对部分解折叠的titin I91的折叠过程进行模拟(图4(c)). 结果发现, 5个AMBER力场产生的折叠中间体和C22*的不同, 含有过多的 α_x 构象, α_x 构象和 α -螺旋构象比较接近, 但是氨基酸残基的 ϕ 二面角更加接近 -180° . AMBER FB15力场还在模拟中破坏了titin I91的稳定部分, 导致不正常的解折叠. 此外, I91折叠中间体的结构和Ace-(AA-QAA)₃-NH₂的平衡模拟表明ff14SB的二面角扭转能垒过高导致螺旋寿命过长.

4 水模型对蛋白质模拟的影响

大多数蛋白质是在水溶液中发挥功能的, 因此对蛋白质进行模拟时, 除了蛋白质力场还需要使用水模型. 常用的水模型有TIP3P^[44]、SPC/E^[63]、TIP4P-Ew^[64]等. 水模型的选择对于有序蛋白质的模拟影响较小, 但是对IDPs和蛋白质非折叠态影响很大. Gupta等^[65]用3种不同的水模型——mTIP3P、TIP4P和TIP4P-Ew对Trp-cage迷你蛋白的折叠态和非折叠态进行模拟, 发现水模型对折叠态没有显著的影响, 而不同水模型产生的非折叠态的螺旋含量和末端距离有很大

图 4 (a) VWF的TIL'E'区域的NMR结构^[51]; (b) 用a99SB-disp模拟产生的p53-TAD的结构^[50]; (c) titin I91的天然态结构(灰色)和在力微扰下的N端结构(蓝色)^[53] (网络版彩图)

Figure 4 (a) Structures of the TIL'E' region of VWF determined by NMR experiments ^[51]; (b) structures derived from a99SB-disp simulations ^[50]; (c) the native structure (gray) and the N-terminal structure under force (blue) of titin I91 ^[53] (color online).

差异, 其中用TIP4P-Ew得到的非折叠态的螺旋含量最低, 末端距离最大。

在分子力场中, 水和蛋白质之间存在L-J相互作用和静电相互作用, 准确描述这些相互作用对蛋白质模拟非常关键。Nayar等^[66]发现使用不同强度L-J相互作用的水模型对Ace-A₅-Nme的模拟结果有很大的影响。增强水模型的L-J相互作用会降低 α -螺旋构象的比例, 提高 β -折叠和PPII构象的比例, 增大多肽的 R_g , 增大多肽-水之间的氢键数目以及降低多肽-多肽和水-水之间的氢键数目。

Best等^[57]最早增加蛋白质和水的L-J相互作用来解决IDPs模拟过于塌缩的问题, 开发了ff03ws和ff99SBws力场。基于同样的思路, Li等^[67]对不同氨基酸原子增加不同程度的蛋白质和水的L-J相互作用来拟合氨基酸侧链类似物的溶剂化自由能, 建立新的力场C36s。C36s缓解了母力场C36^[68]过稳定泛素二聚体的问题, 并且, 用C36模拟EphA2受体的SAM结构域(E-SAM)和SHIP2酶的SAM结构域(S-SAM)的结合过程, 产生的构象中天然复合物构象只占了6.7%, 而C36s能够采样到更多天然复合物构象, 比例达到了16.7%。这些结果说明C36s提高了对蛋白-蛋白结合的模拟。但是, 不改变水-水相互作用, 只增强蛋白-水相互作用的策略可能有一定的局限性。Zhang等^[69]用不同力场和水模型来模拟氨基酸侧链类似物的溶剂化自由能和氨基酸的扩散系数, 发现ff03ws和ff99SBws相比母力场提高了对溶剂化自由能的模拟, 但是对扩散系数的模拟没有改进, 相比实验值仍高估了大约40%。

水的很多独特性质来源于复杂的氢键相互作用, 水模型用静电相互作用来模拟氢键^[70], 不同水模型间静电相互作用的差异对模拟有很大的影响。Anandakrishnan等^[71]用TIP3P、TIP4P-Ew和隐式溶剂模型AMBER igb5对CLN025迷你蛋白进行了折叠模拟, 发现TIP3P和TIP4P-Ew模拟得到的折叠自由能的差值接近2 kcal/mol。为了找出差异的来源, 他们随机挑选了一个折叠结构和一个未折叠结构, 基于这两个结构, 用TIP3P和TIP4P-Ew分别计算了折叠能并且对折叠能进行了分解, 发现两个折叠能的差异主要来源于水-水静电相互作用。Izadi等^[72]用TIP3P、TIP4P-Ew、OPC和隐式溶剂模型GBNSR6计算了15对蛋白质-配体复合物的溶剂化自由能的静电分量(ΔG_{pol})和结合自由能的静电分量($\Delta \Delta G_{\text{pol}}$)。不同水模型的结果有很大的差异,

如TIP3P和TIP4P-Ew的 $\Delta \Delta G_{\text{pol}}$ 的绝对误差最高可以达到~9 kcal/mol。

水动力学和蛋白质动力学之间存在紧密的联系。蛋白质会影响周围水的结构和动力学性质。模拟表明IDPs相比有序蛋白质能够结合更多的水, 周围的水更加规则, 停留时间更长^[73]。水模型同样会影响蛋白质的动力学性质。Salvi等^[74]分别用TIP4P/2005^[75]和TIP3P对125个氨基酸的仙台病毒C端结构域的非折叠态进行模拟, 发现水模型可影响蛋白质链段内的运动, 进而影响对NMR弛豫速率的模拟, TIP4P/2005相比TIP3P能更好地重现NMR弛豫速率, 更准确地描述该蛋白质皮秒到纳秒时间尺度的运动。

传统水模型的开发过程主要优化的是电荷和L-J势能参数, |OH|键长和 $\angle$ HOH键角是固定的或者只允许在实验值附近变化。OPC是2014年发表的通用的四点半水模型^[46]。为了更准确地描述水分子的电荷分布, 模拟液态水的氢键相互作用, OPC取消了传统水模型中|OH|键长和 $\angle$ HOH键角的限制, 拟合了水分子3个最低的多极矩, 对于水分子的11个关键的液相性质, OPC的模拟结果与实验值的平均相对误差为0.76% (图5)。OPC水模型被证明能够改进RNA的模拟结果^[76,77]。不仅如此, OPC搭配C36脂质力场能很好地重现气液界面上脂单层的表面压-分子面积等温线的实验值^[78]。Shabane等^[79]分别用OPC和TIP3P水模型对IDPs和有序蛋白质进行了模拟。对两个IDPs体系的模拟表明, OPC产生了伸展的构象集合, 它的 R_g 更加接近相同氨基酸残基数目的无规卷曲的 R_g , 而TIP3P产生的构象系综非常塌缩, 它的 R_g 更加接近相同残基数目的有序蛋白质的 R_g 。对CLN025迷你蛋白的折叠模拟表明, OPC和TIP3P都能采样到其折叠构象, 但是OPC能够更好地平衡折叠态和非折叠态的比例, 计算得到的折叠自由能和实验值更加接近。

很多力场是在特定水模型中优化得到的, 水模型的缺陷会影响力场对训练集外体系的模拟。很多老的水模型无法很好地重现水的整体性质。例如, TIP3P远远高估了水的自扩散系数, 和实验值的相对误差为139.1% (图5)^[46]。Anandakrishnan等^[71]证明在可加和分子力场中, 溶剂-溶剂相互作用的误差无法通过调整溶质-溶质相互作用和溶剂-溶质相互作用来完全抵消。因此, 开发新力场的时候使用OPC这样更准确的水模型非常重要。

图5 不同水模型对水物理性质的模拟值和实验值的相对误差^[46] (网络版彩图)

Figure 5 Relative error in various properties by different water models^[46] (color online).

5 IDPs结合过程的MD模拟

5.1 IDPs和生物大分子的结合机制

大多数IDPs通过结合DNA或者蛋白质来发挥功能^[80]。当IDPs与它们结合时常常会发生诱导折叠, 形成一个更加有序的结构, 这个过程称为折叠与结合的耦合(coupled folding and binding)^[10,81]。用实验手段可以获得IDPs结合生物大分子的热力学和动力学性质, 如结合自由能和动力学速率常数^[82]。但是, 实验往往只能提供构象系综的平均信息, 不能准确描述IDPs的结合过程。MD能够对实验进行补充, 在原子水平上模拟IDPs和生物大分子的结合, 对结合中间体和关键残

基进行细致的描述^[83]。之前对IDPs结合的研究常常使用粗粒化力场^[84,85]或者隐式溶剂模型^[86]来提高计算效率, 但是这会牺牲一定的准确性。随着图形处理器(graphics processing unit, GPU)^[87]和新的增强抽样方法的出现^[88-90], 越来越多的研究使用全原子力场搭配显式溶剂模型研究IDPs的结合过程(表2)。对于IDPs结合生物大分子的机制, 有两个著名的模型: 诱导契合(induced-fit)^[91]和构象选择(conformational selection)^[92]。诱导契合指的是IDPs和底物的结合诱导IDPs折叠形成最终稳定的复合物, 而构象选择指的是, 游离的IDPs在多种构象间转换, 在采取其中某一种构象时, 和底物结合形成稳定的复合物。

模拟发现很多IDPs结合都遵循诱导契合模型。Collins等^[93]用MD模拟的轨迹来构建马尔科夫状态模型(Markov state model, MSM), 阐明了无序蛋白BrkDBD和DNA完整的结合-折叠途径, 包括各个中间体的结构和时间尺度。MSM显示BrkDBD的绝大多数二级和三级结构是在进入DNA上的结合位点后形成的(图6(a))。Chong等^[94]通过MD和自由能计算构建了磷酸化的激酶诱导结构域(pKID)与受体KIX结合过程的自由能景观图(图6(b))。自由能景观图随着pKID结合KIX变得越来越陡峭, 表明结合有利于pKID的折叠。结合过程中存在一个中间体, 只有pKID的两个 α -螺旋

表2 IDPs结合过程的MD研究

Table 2 MD studies on IDPs' binding process

体系	力场	模拟方法	模拟时长	参考文献
BrkDBD:DNA	ff99SB	MD, MSM	40 μ s MD	[93]
pKID:KIX	C22*	MD	33.5 μ s MD	[94]
N _{TAIL} α -MoRE:XD	a99SB-disp	MD	300 μ s MD	[95]
H1 NTD:DNA	C36m, ff99SB*-ILDN-q	T-REMD, BEMD, PTMetaD-WTE	112 μ s T-REMD, 1.6 μ s BEMD, 22.4 μ s PTMetaD-WTE	[96]
A β :solanezumab	ff99SB-ILDN	McMD, US	41.6 μ s McMD, 19.2 μ s US	[97]
PMI:MDM2	ff99SB-ILDN	MD, H-REMD, MSM	503.6 μ s MD, 100 μ s H-REMD	[98]
p53 CTDF:S100B	ff94和ff96混合力场	V-McMD	6.48 μ s V-McMD	[99]
p53-TAD:MDM2	ff99SB-ILDN	a/dPaCS-MD, MSM	11.675 μ s a/dPaCS-MD	[100]
p53-TAD:MDM2	ff99SB-ILDN	BE-MetaD, InMetaD	27 μ s BE-MetaD	[110]
p53-TAD:EGCG	a99SB-disp, GAFF	REST2	192 μ s REST2	[106]
hIAPP:C60(OH) ₂₄	OPLS	REMD	34.56 μ s REMD	[107]
A β 42:HT, A β 42:SI	CHARMM36m, Charmm GenFF	REST	480 μ s REST	[108]
c-Myc ₄₀₂₋₄₁₂ :10058-F4	ff03*	M&M	5 μ s M&M	[109]

图 6 (a) 用MSM分析得到的BrkDBD的结合-折叠途径^[93]; (b) pKID和KIX结合过程的自由能景观图^[94]; (c) N-TAIL α -MoRE和XD结合过程的自由能景观图^[95] (网络版彩图)

Figure 6 (a) BrkDBD binding-folding pathways from the MSM analysis^[93]; (b) free energy landscape of the binding process between pKID and KIX^[94]; (c) free energy landscape of the binding process between N-TAIL α -MoRE and XD^[95] (color online).

和KIX形成正确的静电相互作用才能形成最终的复合物。Robustelli等^[95]不使用增强抽样方法,在400 K的温度下进行了200 μ s的MD模拟来探究无序蛋白质N-TAIL α -MoRE结合有序蛋白质XD的过程(图6(c))。模拟结果表明,结合中间体的构象非常多样,分析发现结合中间体的形成仅仅依赖4对关键的分子间相互作用。他们还发现存在一种解折叠-结合的机制,高度螺旋的N-TAIL α -MoRE常常需要解折叠来促进关键分子间相互作用的形成,进而形成最终的复合物。Sridhar等^[96]用调和系综中的平行回火元动力学(parallel-tempered metadynamics in the well-tempered ensemble, PTMetaD-WTE)、温度副本交换分子动力学(temperature replica exchange molecular dynamics, T-REMD)和偏交换元动力学(biased-exchange metadynamics, BEMD)对H1连接组蛋白的N端结构域(H1 NTD)的3种亚型进行了研究。模拟发现,3种H1 NTD亚型单独的时候都是完全无序的,但是和DNA结合后会形成不同程度的螺旋。重要的是,不同H1 NTD亚型结合DNA后表现出的螺旋倾

向和其结合核小体的能力密切相关,螺旋倾向越强的亚型结合核小体的能力也越强。

也有底物通过构象选择来结合IDPs。Bekker等^[97]用动态对接和路径抽样模拟发现,处于未结合态和结合中间体时,A β 的结构和solanezumab抗体的互补决定区的结构都会不断变化,只有A β 和互补决定区的结构都处于特定构象时才能稳定的结合。

有一些IDPs结合底物的过程更加复杂,不能用诱导契合或构象选择模型解释。Paul等^[98]用MD、哈密顿量副本交换分子动力学(Hamiltonian replica exchange molecular dynamics H-REMD)和MSM发现无序蛋白质PMI和MDM2存在很多不同的结合路径。绝大多数途径,PMI的结合和折叠同时发生,没有明显的先后次序。Iida等^[99]用虚拟系统耦合多正则分子动力学(virtual-system coupled multicanonical molecular dynamics, V-McMD)发现p53无序的C端结构域的一个片段(CTDf)结合有序蛋白质S100B会形成多种复合物结构。形成复合物的CTDf相比单体的CTDf构象反而更

加多样, 可能是因为游离的p53 CTDf通过分子内氢键和疏水相互作用形成紧凑的结构, 而和S100B结合时形成分子间疏水相互作用, 构象变得更加伸展. 一些增强抽样方法, 如Be-MetaD, 需要设定集体变量(collective variables, CV), CV的选取往往需要复合物天然结构的信息, Tran等^[100]提出了一个不依赖复合物结构信息的增强抽样方法a/dPaCS-MD, 他们用a/dPaCSMD结合MSM, 准确模拟了p53-TAD和MDM2的结合自由能、结合速率常数和解离速率常数. MSM表明, p53-TAD结合MDM2分两步进行, 先通过构象选择形成一个特定的中间体, 再通过诱导契合形成最终的复合物结构.

5.2 小分子结合对IDPs构象系综的影响

很多IDPs在蛋白相互作用网络中作为信号中心和多个靶标存在相互作用^[3]. IDPs的突变或含量变化常和疾病相关, 如癌症、神经退行性疾病、心血管疾病和糖尿病等^[101~104]. 大约1/5疾病相关的错义突变发生在无序区域上^[102]. 越来越多的研究开始以IDPs作为药物靶标, 并且已经发现了一些小分子能够结合IDPs^[12]. 有序蛋白质通常用一个特定的口袋和小分子形成稳定的构象, 而IDPs不同, 一个IDP往往含有多个位点和小分子结合, 形成多种多样的复合物结构. 随着小分子的结合, IDPs可能会形成更加多样的结构, 构象熵增加, 也有可能偏好某一个特定的结构, 构象熵降低^[105].

目前对IDPs和小分子结合的模拟表明, 小分子和IDPs的多个位点都可以结合, 和一些特定的氨基酸结合概率更高, 结合前后IDPs都是无序的. Liu等^[106]用溶剂回火副本交换(replica exchange with solute tempering, REST2)增强抽样技术研究表没食子儿茶素没食子酸酯(epigallocatechin gallate, EGCG)结合对p53-TAD构象系综的影响, 发现EGCG的结合会使得p53-TAD变得塌缩, 并且EGCG和p53-TAD的大多数氨基酸都有一定程度的接触. 遗憾的是, 模拟没有完全收敛, 无法确定哪些氨基酸和EGCG结合得更加频繁. Bai等^[107]用REMD研究富勒醇($C_{60}(OH)_{24}$)抑制人胰岛淀粉样多肽(human islet amyloid polypeptide, hIAPP)聚集的机制. hIAPP二聚体偏好富含 β 片层的结构, 但是随着富勒醇的结合, 二聚体之间的相互作用减少, 富含片层的结构比例大大降低, 塌缩的富含卷曲的结构比例明显增加(图7(a)). 富勒醇主要和极性残基T9、Q10、

N14、N21、N22、N31、N35、T36形成氢键, 以及和疏水残基Y37同时形成氢键和范德华相互作用来结合hIAPP二聚体. 对hIAPP原纤维的MD模拟表明, 富勒醇的结合主要发生在hIAPP淀粉样多肽核心区域, 并且会降低原纤维的稳定性. 牛磺酸(homotaurine, HT)和鲨肌醇(scyllo-inositol, SI)被证实能够抑制A β 聚集, 为了研究它们的抑制机制, Liang等^[108]用REST增强抽样方法对HT和SI与A β 42的结合分别进行了模拟. 两个小分子的结合都会降低A β 42的C端区域的 β -折叠含量, 使得C端区域原本偏好的较伸展的结构变为较塌缩的C形结构(图7(b)). HT带负电的磺酸基主要和A β 42的D1-G9、V24-K28以及G37-G38残基的骨架形成相互作用, 带正电的氨基主要与天冬氨酸和谷氨酸的侧链以及C端的羧基相互作用. 对于SI, 它的多元醇环和A β 42的骨架以及极性残基的侧链都有相互作用. Heller等^[109]用元动力学元推理(metadynamic metainference, M&M), 结合力场和NMR数据, 对致癌基因蛋白c-Myc的402~412片段(c-Myc₄₀₂₋₄₁₂)和小分子10058-F4的结合进行了模拟(图7(c)). 尿素能够和蛋白无差别的结合, 为了验证10058-F4是否特异性地结合c-Myc₄₀₂₋₄₁₂, 他们还对c-Myc₄₀₂₋₄₁₂和尿素的结合进行了模拟. 10058-F4主要结合c-Myc₄₀₂₋₄₁₂的疏水N端, 并且和不同残基的结合倾向有很大差异, 而尿素和所有残基都表现出了类似的结合模式, 这些结果表明c-Myc₄₀₂₋₄₁₂和10058-F4结合是特异性的. 模拟中10058-F4结合最频繁的3个残基是Y402、Q407和L404, 荧光淬灭实验验证了L404P和Q407K突变在一定程度上降低了c-Myc₄₀₂₋₄₁₂和10058-F4的结合能力, 和模拟结果相符合.

6 总结与展望

对于全原子分子动力学模拟来说, IDPs是一类非常具有挑战性的体系. 由于IDPs的不同构象间自由能只有很小的差异, 要准确描述IDPs的构象系综非常困难. 但是, 目前仍然存在一些力场在IDPs体系上有良好的表现, 如a99SB-disp和RSFF2+等. 不仅如此, 新的力场还在被不断发展, 可以期望未来对IDPs的模拟会越来越准确.

水模型在蛋白质模拟中扮演了重要的角色, 水模型的选择对模拟的结果有很大的影响, 尤其是对IDPs, 水模型的选择有时候比力场的选择对模拟结果影响更

图 7 (a) hIAPP二聚体结合富勒醇前后的构象系综中聚类最大类的代表结构^[107]; (b) Aβ42结合HT前后的构象系综^[108]; (c) c-Myc₄₀₂₋₄₁₂与10058-F4结合态的构象系综^[109] (网络版彩图)

Figure 7 (a) The representative conformations for the first most-populated cluster of hIAPP dimer with or without C₆₀(OH)₂₄ [107]; (b) the structural ensembles of Aβ42 with or without HT [108]; (c) the structural ensemble representing the bound state of c-Myc₄₀₂₋₄₁₂ and 10058-F4 [109] (color online).

大. 此外, 力场发展时使用更加准确的水模型可以减轻力场对测试集的过拟合, 使得力场在不同体系上都能有良好的表现.

IDPs通常需要和特定底物结合来发挥功能, MD作为实验的补充, 可以从原子水平上去描述IDPs和底

物的结合过程, 得到实验无法得到的结构和动力学信息. 随着计算效率的提高和增强抽样方法的发展, MD能够研究的体系将越来越大, 能够研究的生物学过程将越来越复杂. 在现在和未来的IDPs研究中, MD都会是一个非常强大的工具.

参考文献

- 1 Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang CH, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC, Obradovic Z. *J Mol Graphics Model*, 2001, 19: 26–59
- 2 van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, Fuxreiter M, Gough J, Gsponer J, Jones DT, Kim PM, Kriwacki RW, Oldfield CJ, Pappu RV, Tompa P, Uversky VN, Wright PE, Babu MM. *Chem Rev*, 2014, 114: 6589–6631
- 3 Wright PE, Dyson HJ. *Nat Rev Mol Cell Biol*, 2015, 16: 18–29
- 4 Shammas SL. *Curr Opin Struct Biol*, 2017, 42: 155–161
- 5 Dyson HJ, Wright PE. *Nat Rev Mol Cell Biol*, 2005, 6: 197–208
- 6 Graewert MA, Svergun DI. *Curr Opin Struct Biol*, 2013, 23: 748–754
- 7 Schuler B, Hofmann H. *Curr Opin Struct Biol*, 2013, 23: 36–47

- 8 Jensen MR, Zweckstetter M, Huang J, Blackledge M. *Chem Rev*, 2014, 114: 6632–6660
- 9 Levine ZA, Larini L, LaPointe NE, Feinstein SC, Shea JE. *Proc Natl Acad Sci USA*, 2015, 112: 2758–2763
- 10 Kasahara K, Terazawa H, Takahashi T, Higo J. *Comput Struct Biotech J*, 2019, 17: 712–720
- 11 Yu C, Niu X, Jin F, Liu Z, Jin C, Lai L. *Sci Rep*, 2016, 6: 22298
- 12 Ruan H, Sun Q, Zhang W, Liu Y, Lai L. *Drug Discovery Today*, 2019, 24: 217–227
- 13 Gibbs EB, Cook EC, Showalter SA. *Archives Biochem Biophys*, 2017, 628: 57–70
- 14 Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. *J Am Chem Soc*, 1995, 117: 5179–5197
- 15 MacKerell AD Jr., Bashford D, Bellott M, Dunbrack Jr. RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. *J Phys Chem B*, 1998, 102: 3586–3616
- 16 Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF. *J Comput Chem*, 2004, 25: 1656–1676
- 17 Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. *J Phys Chem B*, 2001, 105: 6474–6487
- 18 Dauber-Osguthorpe P, Hagler AT. *J Comput Aided Mol Des*, 2019, 33: 133–203
- 19 Hagler AT. *J Comput Aided Mol Des*, 2019, 33: 205–264
- 20 Huang J, MacKerell Jr. AD. *Curr Opin Struct Biol*, 2018, 48: 40–48
- 21 Nerenberg PS, Head-Gordon T. *Curr Opin Struct Biol*, 2018, 49: 129–138
- 22 Jiang F, Zhou CY, Wu YD. *J Phys Chem B*, 2014, 118: 6983–6998
- 23 Zhou CY, Jiang F, Wu YD. *J Phys Chem B*, 2015, 119: 1035–1047
- 24 Kang W, Jiang F, Wu YD. *J Chem Theor Comput*, 2018, 14: 4474–4486
- 25 Yang S, Liu H, Zhang Y, Lu H, Chen H. *J Chem Inf Model*, 2019, 59: 4793–4805
- 26 Song D, Liu H, Luo R, Chen HF. *J Chem Inf Model*, 2020, 60: 2257–2267
- 27 Tian C, Kasavajhala K, Belfon KAA, Raguette L, Huang H, Migués AN, Bickel J, Wang Y, Pincay J, Wu Q, Simmerling C. *J Chem Theor Comput*, 2020, 16: 528–552
- 28 Piana S, Robustelli P, Tan D, Chen S, Shaw DE. *J Chem Theor Comput*, 2020, 16: 2494–2507
- 29 Diem M, Oostenbrink C. *J Chem Inf Model*, 2020, 60: 279–288
- 30 Swindells MB, MacArthur MW, Thornton JM. *Nat Struct Mol Biol*, 1995, 2: 596–603
- 31 Jiang F, Han W, Wu YD. *J Phys Chem B*, 2010, 114: 5840–5850
- 32 Jiang F, Han W, Wu YD. *Phys Chem Chem Phys*, 2013, 15: 3413–3428
- 33 Jiang F, Wu HN, Kang W, Wu YD. *J Chem Theor Comput*, 2019, 15: 2761–2773
- 34 Jiang F, Wu YD. *J Am Chem Soc*, 2014, 136: 9536–9539
- 35 Zhou CY, Jiang F, Wu YD. *J Chem Theor Comput*, 2015, 11: 5473–5480
- 36 Li S, Elcock AH. *J Phys Chem Lett*, 2015, 6: 2127–2133
- 37 Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. *Proteins: Struct, Funct, Bioinf*, 2010, 78: 1950–1958
- 38 Li DW, Brüschweiler R. *Angew Chem Int Ed*, 2010, 49: 6778–6780
- 39 Xun S, Jiang F, Wu YD. *J Chem Theor Comput*, 2015, 11: 1949–1956
- 40 Xun S, Jiang F, Wu YD. *Bioorg Med Chem*, 2016, 24: 4970–4977
- 41 Kang W, Jiang F, Wu YD, Wales DJ. *J Chem Theor Comput*, 2020, 16: 800–810
- 42 Geng H, Jiang F, Wu YD. *J Phys Chem Lett*, 2016, 7: 1805–1810
- 43 Robustelli P, Piana S, Shaw DE. *Proc Natl Acad Sci USA*, 2018, 115: E4758–E4766
- 44 Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. *J Chem Phys*, 1983, 79: 926–935
- 45 Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. *J Chem Theor Comput*, 2015, 11: 3696–3713
- 46 Izadi S, Anandakrishnan R, Onufriev AV. *J Phys Chem Lett*, 2014, 5: 3863–3871
- 47 Reif MM, Winger M, Oostenbrink C. *J Chem Theor Comput*, 2013, 9: 1247–1264
- 48 Carballo-Pacheco M, Ismail AE, Strodel B. *J Chem Theor Comput*, 2018, 14: 6063–6075
- 49 Man VH, He X, Derreumaux P, Ji B, Xie XQ, Nguyen PH, Wang J. *J Chem Theor Comput*, 2019, 15: 1440–1452

- 50 Liu X, Chen J. *J Chem Theor Comput*, 2019, 15: 4708–4720
- 51 Kuzmanic A, Pritchard RB, Hansen DF, Gervasio FL. *J Phys Chem Lett*, 2019, 10: 1928–1934
- 52 Petrovic D, Wang X, Strodel B. *Proteins: Struct, Funct, Bioinf*, 2018, 86: 935–944
- 53 Wang D, Marszalek PE. *J Chem Theor Comput*, 2020, 16: 3240–3252
- 54 Rauscher S, Gapsys V, Gajda MJ, Zweckstetter M, de Groot BL, Grubmüller H. *J Chem Theor Comput*, 2015, 11: 5513–5524
- 55 Henriques J, Skepö M. *J Chem Theor Comput*, 2016, 12: 3407–3415
- 56 Ouyang Y, Zhao L, Zhang Z. *Phys Chem Chem Phys*, 2018, 20: 8676–8684
- 57 Best RB, Zheng W, Mittal J. *J Chem Theor Comput*, 2014, 10: 5113–5124
- 58 Gerben SR, Lemkul JA, Brown AM, Bevan DR. *J Biomol Struct Dyn*, 2014, 32: 1817–1832
- 59 Shiltagh N, Kirkpatrick J, Cabrita LD, McKinnon TAJ, Thalassinou K, Tuddenham EGD, Hansen DF. *Blood*, 2014, 123: 4143–4151
- 60 Wu HN, Jiang F, Wu YD. *J Phys Chem Lett*, 2017, 8: 3199–3205
- 61 Marszalek PE, Lu H, Li H, Carrion-Vazquez M, Oberhauser AF, Schulten K, Fernandez JM. *Nature*, 1999, 402: 100–103
- 62 Piana S, Lindorff-Larsen K, Shaw DE. *Biophys J*, 2011, 100: L47–L49
- 63 Berendsen HJC, Grigera JR, Straatsma TP. *J Phys Chem*, 1987, 91: 6269–6271
- 64 Horn HW, Swope WC, Pitera JW, Madura JD, Dick TJ, Hura GL, Head-Gordon T. *J Chem Phys*, 2004, 120: 9665–9678
- 65 Gupta M, Nayar D, Chakravarty C, Bandyopadhyay S. *Phys Chem Chem Phys*, 2016, 18: 32796–32813
- 66 Nayar D, Chakravarty C. *J Phys Chem B*, 2015, 119: 11106–11120
- 67 Li Z, Buck M. *J Chem Theor Comput*, 2019, 15: 4318–4331
- 68 Huang J, MacKerell Jr AD. *J Comput Chem*, 2013, 34: 2135–2145
- 69 Zhang H, Yin C, Jiang Y, van der Spoel D. *J Chem Inf Model*, 2018, 58: 1037–1052
- 70 Onufriev AV, Izadi S. *WIREs Comput Mol Sci*, 2018, 8: e1347
- 71 Anandakrishnan R, Izadi S, Onufriev AV. *J Chem Theor Comput*, 2019, 15: 625–636
- 72 Izadi S, Aguilar B, Onufriev AV. *J Chem Theor Comput*, 2015, 11: 4450–4459
- 73 Rani P, Biswas P. *J Phys Chem B*, 2015, 119: 10858–10867
- 74 Salvi N, Abyzov A, Blackledge M. *Sci Adv*, 2019, 5: eaax2348
- 75 Abascal JLF, Vega C. *J Chem Phys*, 2005, 123: 234505
- 76 Bergonzo C, Cheatham III TE. *J Chem Theor Comput*, 2015, 11: 3969–3972
- 77 Yang C, Lim M, Kim E, Pak Y. *J Chem Theor Comput*, 2017, 13: 395–399
- 78 Javanainen M, Lamberg A, Cwiklik L, Vattulainen I, Ollila OHS. *Langmuir*, 2018, 34: 2565–2572
- 79 Shabane PS, Izadi S, Onufriev AV. *J Chem Theor Comput*, 2019, 15: 2620–2634
- 80 Sugase K, Dyson HJ, Wright PE. *Nature*, 2007, 447: 1021–1025
- 81 Huang YQ, Liu ZR. *Acta Physico-Chim Sin*, 2010, 26: 2061–2072
- 82 Yadahalli S, Neira JL, Johnson CM, Tan YS, Rowling PJE, Chattopadhyay A, Verma CS, Itzhaki LS. *Sci Rep*, 2019, 9: 693
- 83 Baker CM, Best RB. *WIREs Comput Mol Sci*, 2014, 4: 182–198
- 84 Knott M, Best RB. *J Chem Phys*, 2014, 140: 175102
- 85 Xiong J, Gao M, Zhou J, Liu S, Su Z, Liu Z, Huang Y. *Proteins: Struct, Funct, Bioinf*, 2019, 87: 265–275
- 86 Sun B, Cook EC, Creamer TP, Kekenos-Huskey PM. *Biochim Biophys Acta*, 2018, 1862: 2651–2659
- 87 Anderson JA, Lorenz CD, Travesset A. *J Comput Phys*, 2008, 227: 5342–5359
- 88 Higo J, Umezawa K, Nakamura H. *J Chem Phys*, 2013, 138: 184106
- 89 Tiwary P, Parrinello M. *Phys Rev Lett*, 2013, 111: 230602
- 90 Bonomi M, Parrinello M. *Phys Rev Lett*, 2010, 104: 190601
- 91 Koshland DE. *Angew Chem Int Ed Engl*, 1994, 33: 2375–2378
- 92 Tsai CJ, Ma B, Nussinov R. *Proc Natl Acad Sci USA*, 1999, 96: 9970–9972
- 93 Collins AP, Anderson PC. *Biochemistry*, 2018, 57: 4404–4420
- 94 Chong SH, Im H, Ham S. *ACS Cent Sci*, 2019, 5: 1342–1351
- 95 Robustelli P, Piana S, Shaw DE. *J Am Chem Soc*, 2020, 142: 11092–11101

- 96 Sridhar A, Orozco M, Collepardo-Guevara R. *Nucl Acids Res*, 2020, 48: 5318–5331
- 97 Bekker GJ, Fukuda I, Higo J, Kamiya N. *Sci Rep*, 2020, 10: 1406
- 98 Paul F, Noé F, Weikl TR. *J Phys Chem B*, 2018, 122: 5649–5656
- 99 Iida S, Kawabata T, Kasahara K, Nakamura H, Higo J. *J Chem Theor Comput*, 2019, 15: 2597–2607
- 100 Tran DP, Kitao A. *J Chem Theor Comput*, 2020, 16: 2835–2845
- 101 Lu H-C, Chung SS, Fornili A, Fraternali F. *Front Mol Biosci*, 2015, 2: 47
- 102 Vacic V, Markwick PRL, Oldfield CJ, Zhao X, Haynes C, Uversky VN, Iakoucheva LM. *PLoS Comput Biol*, 2012, 8: e1002709
- 103 Uversky VN, Davé V, Iakoucheva LM, Malaney P, Metallo SJ, Pathak RR, Joerger AC. *Chem Rev*, 2014, 114: 6844–6879
- 104 Zhou J-B, Xiong Y, An K, Ye Z-Q, Wu Y-D. *Bioinformatics*, 2020
- 105 Heller GT, Bonomi M, Vendruscolo M. *J Mol Biol*, 2018, 430: 2288–2292
- 106 Liu X, Chen J. *Pac Symp Biocomput*, 2020, 25: 195–206
- 107 Bai C, Lin D, Mo Y, Lei J, Sun Y, Xie L, Yang X, Wei G. *Phys Chem Chem Phys*, 2019, 21: 4022–4031
- 108 Liang C, Savinov SN, Fejzo J, Eyles SJ, Chen J. *J Chem Theor Comput*, 2019, 15: 5169–5174
- 109 Heller GT, Aprile FA, Bonomi M, Camilloni C, De Simone A, Vendruscolo M. *J Mol Biol*, 2017, 429: 2772–2779
- 110 Zou R, Zhou Y, Wang Y, Kuang G, Ågren H, Wu J, Tu Y. *J Chem Inf Model*, 2020, 60: 1551–1558

Recent development of atomistic force fields and simulations of intrinsically disordered proteins

Jianan Chen², Yundong Wu^{1,2,3*}

¹ Shenzhen Bay Laboratory, Shenzhen 518055, China

² Peking University Shenzhen Graduate School, Shenzhen 518055, China

³ College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China

*Corresponding author (email: wuyd@pkusz.edu.cn)

Abstract: Intrinsically disordered proteins (IDPs) are a special class of proteins which lack well-defined structures under native conditions while performing essential biological functions. Molecular dynamics (MD) simulation is a powerful tool to unravel the atomistic details on IDPs' structures and dynamics, which cannot be obtained from experiments. In this review, we first review the recent development of protein force fields and some benchmark studies. Then, we discuss the effect of water model choice on protein simulations. Finally, we introduce some MD applications in studying IDPs' binding process.

Keywords: molecular dynamics simulations, protein force field, water model, intrinsically disordered proteins

doi: [10.1360/SSC-2020-0143](https://doi.org/10.1360/SSC-2020-0143)